

## Chapter Thirteen

### GRAMMATICAL COMPETENCE OF IFA UL STUDENTS\*

#### INTRODUCTION

Each year some 50-60 new students are admitted to IFA UL (Institute of English Studies, University of Łódź) to study linguistics and literature. Admission is by entrance examination and successful candidates constitute ca. 40% of the applicants. It is usual for the candidates to have completed a basic 4-year (high school) course in English and not unusual for this to have been supplemented by some additional experience in English. In addition to courses in theoretical and applied linguistics, British and American literature, and British and American history and culture, all taught in English, the students all take a set of English language courses (practical phonetics, comprehension, composition, conversational English, grammar, and translation from and into English). During the ten fifteen-week terms at the Institute the students take an average of 10-12 hours a week of literature, linguistics and history courses and 9 hours of language work. Then there are the videos they watch, the tapes they listen to, the essays they write and the books they read as homework or extracurricular activities which, taken together with the coursework, represents quite a substantial amount of exposure to the language.

The progress the students make in English is assessed at the end of each academic year, but the English exams are designed

---

\* Jerzy Tomaszczyk, University of Łódź.

based on the actual work each year-group has done during the past year and they are thus meant -- naturally enough -- to establish the extent to which a certain prescribed amount of language knowledge or skill has been acquired, allowing little opportunity for direct comparisons among the different year-groups. Consequently, whatever knowledge the teaching staff have of the students' progress in global English proficiency from one year of studies to the next can only be highly impressionistic. One such impression that has often been voiced is that the students' overall proficiency in English increases up to about the third year following which it begins to decline, and at least on one occasion the examination committee were agreed that many of the fourth year people would not have passed the third year English exam.

The idea behind the study reported here -- based on a single test administered to ca. 50% of the entire student body -- was to see how the different year-groups would do on one and the same test. More specifically, the study was expected to throw some light on how the subjects' language competence develops during the five years they spend at the university, how each year-group compares with the others, and how they all compare with native speakers of British English.

The instrument used for the purpose was a syntactic acceptability test adopted from Strässler [1984]. While such tests can only probe one aspect of learners' proficiency, it is not an unimportant one under conditions of (adult) FL learning, the knowledge tested being of the kind that is drawn upon e.g. when one assesses somebody else's spoken or written performance or before one hands in a piece of one's own written work. One advantage of acceptability tests is that they are easy to administer to a large and varied group of subjects, and the results are easy to analyze. The one used in the present study had the added advantage of having already been done by a group of 18 native speakers of British English (referred to below as the NS group). The results could thus be used as a standard against which to compare our students' responses.

## PROCEDURE

Each student was presented with an identical set of 26 sentences (Table 1) and asked to judge them as to their acceptability. The options suggested to the subjects were: *fully acceptable* (grammatically correct) (+); *unacceptable* (grammatically incorrect) (-); *marginally acceptable* (borderline case) (?); *don't know* ( $\emptyset$ ). The *don't know* option was introduced so that possible cases of lack of understanding did not affect the other judgements. Since the students had all heard about the acceptability-grammaticality distinction, they were told to focus on the former. The phrasing 'acceptable = grammatical' was nevertheless retained because such is the popular understanding of the terms and because that is how the instructions were phrased in the native speaker test (Strässler's study).

Table 1  
Test sentences and native speaker responses  
(% rel. to  $n = 18$ )

Test sentences	+	-	?
1. The other team has advantage.		38.9	61.1
2. Their experience was taken advantage by everyone.		100	
3. Last Saturday England won the game, didn't it?		100	
4. Last Saturday England won the game, didn't they?	100		
5. Some dirt was under the rug.	33.3		66.6
6. There was some dirt under the rug.	100		
7. The policeman was here today.	100		
8. There was the policeman here today..	27.8	44.4	27.8
9. Dick insisted himself to be fearless.		100	
10. Dick proclaimed himself to be fearless.	100		
11. Susan believes to win the race.		100	
12. That the world is round was proved by Magellan.	100		
13. It was proved by Magellan that the world is round.	100		

Test sentences	+	-	?
14. John said that it disturbed Sue to make a fool of himself in public.		100	
15. That washing himself with mud disturbed Sue, surprised Pete.	44.4	27.8	27.8
16. He is a potential candidate.	100		
17. He is a candidate who is potential.		100	
18. Though that Mary is pretty is clear, I still think Alice is prettier.	33.3	66.6	
19. Pretty though it is clear that Mary is, I still think Alice is prettier.		72.2	27.8
20. For Tom to win is certain.		77.8	22.2
21. That that pineapples grow on pinetrees is true is unlikely.		100	
22. It is unlikely that that pineapples grow on pinetrees is true.		100	
23. Mark is exciting to talk to about politics.	100		
24. There are believed by the police chief to be many criminals still at large.	27.8	27.8	44.4
25. I expected there to be a man behind the counter.	100		
26. You re being silly, aren't you.	100		

The test was conducted under classroom conditions with classes of 5-15 students. The test papers, which also included a personal information and language learning history section, took ca. 10 min. to complete. The tests were administered in March and April 1988, i.e., in the middle of the spring semester.

The figures next to each sentence in Table 1 represent percentages of native speaker responses. For example, the 100 in the '-' column next to sentence No. 2 means that all British subjects considered that sentence unacceptable. If we treat these figures not as percentages but as points on a 1-100 scale, we can add them all up -- one from each row -- and get one figure representing native speaker competence. Since there are a number of figures next to some of the sentences, I decided to add up the higher figures from each row to get the upper limit of NS

competence. That figure is 2277.5 and it is used in this study as the standard (100%) against which all the student responses are measured and in terms of which they are compared. It is a complex figure being made up of the following components: 1000 for the 10 sentences unanimously declared 'acceptable', 800 for the 8 sentences unanimously declared unacceptable, and 477.5 for the remaining 8 sentences (those on which the NS's disagreed). It is easy to see that the lower figures in the 'divided opinion' rows and the two other blocks of figures (1000 and 800) add up to 2038.9 and this figure represents the lower bound of the NS responses (89.52% of the upper limit). Individual NS responses were not available to me but it is obvious that no NS overall score could have fallen outside that range.

Every single student response was assigned a numerical value corresponding to an appropriate figure in Table 1. Thus, if a student found sentence 1 fully acceptable, s/he scored nothing; if s/he found it unacceptable, the value assigned to that response was 38.9; and if s/he found it marginally acceptable, the value assigned was 61.1. No values were assigned to 'don't know' decisions. The set of 26 values for each individual was then added up and the total represented an overall score which was converted -- for convenience -- into a % value (relative to 2277.5). The scores thus obtained were analysed using the IFASTATS package developed by D. Coleman (cf. Chapter 5).

## RESULTS

The overall scores for the five year-groups are presented in Table 2.

As can be seen, the mean scores for the five groups confirm the impressionistic observation that the students' competence increases up to the third year of studies following which it gradually declines. Note that the characteristic produced by the mean scores is not really linear, the increase between groups 2 and 3 and the decline between groups 3 and 4 being much more distinct than those between groups 1 and 2, and 4 and 5 respectively. Also the decline between 4 and 5 is less pronounced than

Overall scores for groups 1-5

Group	n	Mean score	SD	Median	Ranges of individual scores (%)
1	28	57.13	5.92	58.90	20.0 (44.88-64.88)
2	30	59.87	11.4	63.42	44.88 (29.03-73.91)
3	28	67.37	9.93	68.9	36.80 (47.81-84.61)
4	27	61.48	9.22	63.42	33.42 (43.17-76.59)
5	21	59.92	6.95	60.54	24.63 (45.86-70.49)
1-5	134	61.19			

the increase between 1 and 2. It has to be emphasized, in this connection, that the design employed was cross-sectional (five groups at different stages of development) and not longitudinal (one group studied at regular time intervals). It follows that if a similar test were to be administered to the same population in the Spring of 1989, what were the second year people in 1988 might not necessarily do as well on it as the third year people did in 1988, etc. While the (maximum possible) range of individual NS scores was just over 10%, it is from two to over four times as wide in the student groups.

To find out whether the differences between the mean scores are significant, the data were subjected to statistical treatment using the Scheffé test, a common and conservative method of multiple post hoc comparison [cf. Hatch and Farhady, 1982: 143-146].

To test the significance of the comparisons one needs to calculate the  $t$  critical value, represented in this case as  $t'$  (to distinguish it from the  $t$  critical value of the  $t$ -test) and the  $t$  observed values for each pair of means. The former value is calculated from the formula

$$(1) t'_{\text{crit}} = \sqrt{(K - 1) F_{\text{crit}} (\alpha, \text{d.f.B}, \text{d.f.W})},$$

where  $K$  represents the number of groups, and  $F$  is the critical  $F$  value for the between and within group degrees of freedom ( $\text{d.f.B}$  and  $\text{d.f.W}$ ) at the selected level of significance ( $\alpha$ ). In the ca-

se under study K is 5, d.f.B is 4 (5 groups) and d.f.W is 129 (134 subjects;  $N - K = 129$ ). Setting  $\alpha$  at .05, the F crit for the intersect of these degrees of freedom in the F distribution table is 2.43. Consequently,

$$\begin{aligned} t'_{\text{crit}} &= \sqrt{(5 - 1) 2.43} \\ &= \sqrt{(4) (2.43)} \\ &= \sqrt{9.72} \\ &= 3.12 \end{aligned}$$

The formula to use for calculating the t observed values is

$$(2) (a) \quad t_{\text{obs}} = \frac{\hat{c}}{\sqrt{2MSW+n}}$$

if the groups under comparison are of equal size, or

$$(2) (b_1) \quad t_{\text{obs}} = \frac{\hat{c}}{\sqrt{\frac{MSW}{n_1} + \frac{MSW}{n_2}}}, \text{ or}$$

$$(b_2) \quad t_{\text{obs}} = \frac{\hat{c}}{\sqrt{MSW \left( \frac{1}{n_1} + \frac{1}{n_2} \right)}}$$

if they are not.  $\hat{c}$  is the estimated comparison for the population and is found from the formula

$$(3) \quad \hat{c} = w_1 \bar{X}_1 + w_2 \bar{X}_2 + \dots + w_K \bar{X}_K$$

where  $w$  is the weight assigned to the given group, and  $\bar{X}$  is the mean score for that group. In practical terms the calculation reduces to subtracting the lower mean score from the higher one e.g. for groups 1 and 2,

$$\begin{aligned} \hat{c}_1 &= (-1) (57.13) + (+1) (59.76) \\ &= -57.13 + 59.76 \\ &= 2.63 \end{aligned}$$

The MSW (within group variance) for the data under discussion is obtained from the IFAMEANS programme via option 5 (COMPARE COLUMN AVERAGES) and via option 3 of the COMPARE COLUMN AVERAGES menu if the word ALL is typed. In the case under study the MSW is 81.84. Thus,

$$\begin{aligned}
 t_{\text{obs}} (\text{Group 1 vs. Group 2}) &= \frac{2.63}{\sqrt{81.84 \left( \frac{1}{28} + \frac{1}{30} \right)}} \\
 &= \frac{2.63}{\sqrt{5.65}} \\
 &= \frac{2.63}{2.38} \\
 &= 1.11
 \end{aligned}$$

As can be seen, the  $t_{\text{obs}}$  for groups 1 and 2 is smaller than the  $t_{\text{crit}}$  which means that the difference between the mean scores for these groups is statistically nonsignificant (at the selected significance level).

The above procedure was applied to all possible combinations of the mean scores yielding the following table (Table 3):

Table 3

Scheffé test for differences in syntactic  
acceptability scores (overall means)  
for the five year-groups

Group	Group 1	Group 2	Group 3	Group 4	Group 5
	$\bar{X} = 57.13$	$\bar{X} = 59.87$	$\bar{X} = 67.37$	$\bar{X} = 61.48$	$\bar{X} = 59.92$
1	-	1.11	4.23**	1.78	1.01
2	-	-	2.87	.67	.02
3	-	-	-	2.42	2.82
4	-	-	-	-	.59
5	-	-	-	-	-

\*\*  $p < 0.5$

As can be seen from Table 3, of the ten comparisons only one  $t_{\text{obs}}$  reveals a significant difference (group 1 vs. group 3) marked with '\*\*' while in two other cases (2 vs. 3 and 3 vs. 5) the differences are fairly close to being significant (Scheffé's is a conservative test).

While it was impossible for me to include the native speaker scores in the comparisons (having no access to individual results), both the magnitude of the difference between the highest scoring group (3) and the native subjects' score (somewhere in



the 89.52-100 range) and the obviously large difference in the range of individual scores (within group variance) between them suggest a high level of significance. However, as has already been indicated, what we are interested in here is the group by group picture of IFA students' competence and the native speaker data are used only as a frame of reference.

The predominantly non-significant nature of the differences between the scores for the five year-groups cannot be interpreted in any straightforward manner but, in view of the limitations of the study hinted at in the introductory section, the reader would be well-advised not to jump to conclusions and decide that no, or very little, acquisition takes place during the five years the students spend at the Institute.

Rather than speculate about the possible causes or implications of the results presented thus far, let us try to see what other information can be extracted from the scores. That there might be more to them than meets the eye at first sight is suggested by the observation that the lower limits of the group ranges form an approximately flat -- straight and horizontal -- characteristic (with the low score for the second year falling out of line), while the upper limits form a curve that is very similar to that produced by the means (cf. Table 2).

To make easier the rather tedious manual handling of the test papers I performed calculations on the raw data by stages, adding up the individual scores by blocks of figures. This procedure led to a number of impressionistic observations which suggested that comparisons might be made to see

(4) how the students did on the sentences on which the native speakers were unanimous (individual scores of 100 in each case);

(a) how they did on the sentences which the native speakers all found fully acceptable (100+);

(b) how they did on the sentences which the native speakers all found fully unacceptable (100-);

(5) how they did on the sentences on which the native speakers disagreed (+, -, ?);

(6) how they did on the sentences which the native speakers judged marginally acceptable (?);

(7) how they used the (+), (-), (?), and (ø) options regard-

less of whether their decisions were correct or incorrect (when compared with NS decisions);

(8) how unanimous they were - the entire population and group by group - on the individual sentences.

The results are presented below in Tables 4-8.

Table 4a

Student scores on the sentences which the NS's unanimously accepted or rejected  
(calculated as % relative to 1800)

Group	n	Mean score	SD	Range	Median
1	28	52.86	10.24	38.89-66.67	55.56
2	30	58.89	12.17	27.78-77.78	66.67
3	28	69.25	12.23	50.0-94.44	72.22
4	27	61.73	11.46	38.89-83.33	61.11
5	21	62.44	8.94	44.44-77.78	61.11
1-5	134	60.92			

Scheffé test:  $t_{\text{obs}}$  (1 vs. 3) = 5.48\*\* and  $t_{\text{obs}}$  (2 vs. 3) = 3.52\*\*;  $t_{\text{obs}}$  (1 vs. 5) = 2.97, nonsignificant but close to being significant; other differences nonsignificant.

\*\*  
p < .05

Table 4b

Student scores on the sentences which the NS's  
unanimously accepted (% rel. to 1000)

Group	n	Mean score	SD	Range	Median
1	28	57.88	10.31	40-80	60
2	30	58.67	15.25	20-80	60
3	28	80.36	12.32	60-100	80
4	27	61.96	17.28	30-100	60
5	21	67.14	13.09	40-90	70
1-5	134	65.22			

Scheffé test: 1 vs. 3 = 6.11\*\*, 2 vs. 3 = 5.94\*\*, 3 vs. 4 = 4.72\*\*, 3 vs. 5 = 3.33\*\*, other differences nonsignificant.

\*\*  
p < .05

Table 4c

Student scores on the sentences which the NS's  
unanimously rejected (% rel. to 800)

Group	n	Mean score	SD	Range	Median
1	28	50.0	16.67	12.5-75.0	50
2	30	59.17	18.84	12.5-87.5	62.5
3	28	55.36	22.16	12.5-87.5	50
4	27	61.11	20.32	25.0-100	62.5
5	21	56.55	17.51	37.5-100	50
1-5	134	56.44			

Scheffé test: no significant differences.

Table 5

Student scores on the sentences on which the NS's  
disagreed (% rel. to 477.5)

Group	n	Mean score	SD	Range	Median
1	28	66.05	12.26	45.38-93.05	66.89
2	30	63.05	18.56	32.57-93.15	64.56
3	28	60.77	11.97	41.89-88.4	60.66
4	27	58.97	15.77	25.59-93.05	58.16
5	21	50.47	13.7	27.92-69.80	52.31
1-5	134	60.41			

Scheffé test:  $t_{\text{obs}} (1 \text{ vs. } 5) = 3.65^{**}$ ;  $t_{\text{obs}} (2 \text{ vs. } 5) = 2.99$ , nonsignificant, but close to being significant.

$^{**} p < 0.5$

Of the many points that could be made about the data presented in Tables 4-6, the following seem to be especially interesting.

Relatively speaking, the greatest contribution to the overall scores comes from decisions related to sentences unanimously accepted by the NS's and those on which the NS's disagreed. Also

Student scores on the sentences which the NS's  
found marginally acceptable (% rel. to 277.7)

Group	n	Mean score	SD	Range	Number of S's with (6) scores
1	28	20.64	17.67	0-57.98	8 (28.6%)
2	30	22.13	15.63	0-49.98	6 (20%)
3	28	25.93	15.1	0-60.0	1 (3.6%)
4	27	19.70	15.41	0-53.98	4 (14.8%)
5	21	16.85	17.91	0-53.98	9 (42.9%)
1-5	134	21.29			

Scheffé test: no significant differences.

in terms of the number of significant differences involved do these two aspects of competence appear to be the more important (discriminating).

That the students did so much better on 'positive' knowledge than on 'negative' knowledge was only to be expected, but that they did as well as they did on 'equivocal' knowledge is somewhat surprising, considering how strange some of the test sentences are. However, note the trend in mean scores for groups 1 to 5 in Table 5.

Perhaps even more interesting is the fact that the contribution from the decisions related to marginally acceptable sentences is as small as it is and that it declines in the higher year-groups. The number (proportion) of the people who scored nothing here produces a characteristic that is approximately a mirror image of the one made by the mean scores. In most cases the medians are quite close to the means, suggesting that close-to-normal distributions are involved.

Most significantly perhaps, while the NS score on the '?' decisions constitutes from 12.2 to 13.6% of the overall score (depending on whether it is calculated relative to the upper or lower limit respectively), the student scores on that option represent on the average, some 3.9% of their overall scores, the range being 0-7%.

Table 7a

The use of the (+) option (% occurrences rel. to  $n \times 26$ )

Group	n	$n \times 26$	Group mean	SD	Range
1	28	728	36.81	10.28	15.38-61.54
2	30	780	32.95	9.45	11.54-53.85
3	28	728	43.41	9.17	26.92-61.54
4	27	702	38.66	13.29	23.08-73.08
5	21	546	42.9	12.43	19.23-65.38
1-5	134	3484	38.59		

Scheffé test:  $t_{\text{obs}}$  (2 vs. 3) = 3.64\*\*;  $t_{\text{obs}}$  (2 vs. 5) = 3.20\*\*.\*\*  $p < 0.5$ 

Table 7b

The use of the '-' option (% occurrences rel. to  $n \times 26$ )

Group	n	$n \times 26$	Group mean	SD	Range
1	28	728	43.82	9.03	19.23-57.69
2	30	780	44.74	12.43	15.38-73.08
3	28	728	35.17	9.4	11.54-57.69
4	27	702	40.73	13.1	15.38-69.23
5	21	546	38.22	11.54	19.23-69.23
1-5	134	3484	40.72		

Scheffé test:  $t_{\text{obs}}$  (2 vs. 3) = 3.25\*\*.\*\*  $p < 0.5$ 

Let us now look at how the different year-groups used the four options (+, -, ?,  $\emptyset$ ). The means etc. presented in Tables 7a-d are derived from cumulative figures resulting from simple addition of the pluses, minuses, etc. as they appeared in the test papers. The percentages are calculated relative to the total number of options available to each group, i.e.,  $n \times 26$  (26 being the number of test sentences).

A total of three S's (two in group 1 and one in group 2) did not take advantage of the '?' option and only 49 (ca. 37%) of

Table 7c

The use of the (?) option (% occurrences rel. to n x 26)

Group	n	n x 26	Group mean	SD	Range
1	28	728	17.17	8.4	0-34.62
2	30	780	18.08	7.08	0-30.77
3	28	728	17.99	6.46	3.85-30.77
4	27	702	18.11	9.41	3.85-42.31
5	21	546	14.47	10.87	0-34.62
1-5	134	3484	17.31		

Scheffé test:  $t_{\text{obs}}$  (4 vs. 5) = 1.49, i.e. far from being significant.

Table 7d

The use of the (ϕ) (don't know) option (% occurrences rel. to n x 26)

Group	n	n x 26	Group mean	SD	Range
1	28	728	2.2	3.21	0-11.54
2	30	780	4.23	7.98	0-30.77
3	28	728	3.43	4.95	0-15.38
4	27	702	2.56	4.53	0-19.23
5	21	546	4.95	7.99	0-23.08
1-5	134	3484	3.42		

Scheffé test:  $t_{\text{obs}}$  (1 vs. 5) = 1.59, i.e. far from being significant.

Table 7e

The use of the (?) and (ϕ) options (combined data)  
(% occurrences rel. to n x 26)

Group	n	n x 26	Group mean	SD	Range
1	28	728	18.54	8.95	3.85-34.62
2	30	780	22.4	9.82	0-42.31
3	28	728	21.91	8.47	3.85-34.62
4	27	702	20.44	10.31	7.69-42.31
5	21	546	19.66	12.39	0-34.62
1-5	134	3484	20.67		

Scheffé test:  $t_{\text{obs}}$  (1 vs. 2) = 1.48, i.e. far from being significant.

all S's used the 'don't know' option (11 in group 1, 10 in 2, 11 in 3, 10 in 4 and 7 in 5, the proportions ranging from .33 to .39). Interestingly, of the ones who did use it, those in groups 5 and 2 used it about twice as much as those in group 1.

Since the 'don't know' option was not available to the NS's, we can combine the data for the (?) and (ø) options, thus assuming that they have the same meaning. As has already been pointed out, the (ø) option was introduced to reduce the amount of guesswork.

One subject in group 2 and three S's in group 5 never used either the (?) or the *don't know* option. While none of the differences between the means is significant and while the trend is for their values to gradually decline starting from group 2, they are nevertheless all quite close to the overall average for the entire population. Note also how close their means are to the mean scores on the sentences which the NS's found marginally acceptable (the *proper* use of the (?) option). It is just possible that there is something non-language specific to the amount of doubt one has about bits and pieces of language but, as should be clear from the data in Table 6 and the last of the comments that follow that table, being appropriately uncertain in a foreign language may well be one of the most difficult things to learn.

What is noticeable about the data listed in Tables 7a-7d is the absence of any patterns in the mean scores, particularly if one compares the group means for the use of the (+) and (-) options. While the NS's used those two options to almost exactly the same extent, the student means produce see-saw patterns that are out of sync.

Finally, let us take a quick look at the scores on the individual sentences to determine the extent of unanimity for the different groups. In keeping with what was said in the comments relative to Tables 5 and 6 about positive knowledge appearing to be easier than negative knowledge, the highest level of inter-subject agreement (the entire population) was on sentences unanimously accepted by the NS's. These are sentences No. 6 (95.52% of the entire population found it acceptable), No. 16 (94.70%), No. 13 (93.28%) and No. 7 (84.33%). However, the 17 sentences on which the level of agreement was 50% or more include 7 (+)

sentences, 7 (-) sentences, and 3 sentences on which the NS's disagreed.

To obtain single figures representing the extent of each group's unanimity on all sentences taken together, I added up -- for each sentence -- the differences between each group's most frequent and least frequent decisions (expressed as percentages) and divided the totals by the number of sentences. The reasoning was as follows. If all the subjects in a group agreed that a particular sentence was e.g. acceptable, the individual scores would always be 100 (points) and the mean score for the group on that sentence would be 100 (%). The extent of unanimity would then be 100 ( $100 - 0 = 100$ ). Such was indeed the case twice in group 3 (sentences 6 and 17) and twice in group 5 (sentences 6 and 13). On the other hand, if a group was equally divided in their decisions (25% (+), 25% (-), 25% (?) and 25% ( $\emptyset$ )), the extent of unanimity would be zero ( $25 - 25 = 0$ ), a hypothetical situation. The results are listed in Table 8.

Table 8

The extent of group unanimity on all sentences

Group	n	Mean difference	SD	Range of differences
1	28	62.47	19.19	21.42-96.43
2	30	59.60	19.84	26.67-96.67
3	28	63.5	23.48	29.72-100
4	27	59.32	20.35	22.22-96.3
5	21	62.05	23.64	23.81-100

Scheffé test:  $t_{\text{obs}}$  (3 vs. 4) = 0.71, i.e. very far from being significant.

Note, that -- strictly speaking --  $t_{\text{crit}}$  has a slightly different value than before and, likewise, the  $t_{\text{obs}}$  is calculated using the simpler version of the relevant formula, i.e., (2a). The reason is that in this case the columns are each made up of figures relating to the 26 sentences, which means that the number of the d.f.W is now 125 and, consequently, the  $F_{\text{crit}}$  value to substitute in formula (1) is 2.44, yielding a  $t_{\text{crit}}$  value of 3.124, while the one previously used was 3.117. In practice, it



is 3.12 in both cases as a result of rounding. The difference is obviously insignificant for our purpose and the whole thing is mentioned at all only as a matter of principle.

#### DISCUSSION

Owing to the nature and the relatively small size of the test employed as well as the nature of the design (cross-sectional), whatever conclusions one would want to draw from the results presented above would necessarily be highly tentative. One thing that seems fairly certain, though, is that FL competence differs from native speaker competence not only in quantitative but also in qualitative terms. In other words, FL learners not only know less but they also know what they do know somewhat differently. It would certainly be very useful to extend the study by including groups of beginning and early intermediate EFL learners as well as at least one group of people representing a higher level of English proficiency than the IFA students. In this way it might be possible to find out how the trends noted here originate and where they go next. Altogether a total of 120 comparisons were made ( $12 \times 10$ ) of which 11 were statistically significant and another 4 were close to being significant. This, together with the fact that there is a lot of individual variation within the student groups whichever way one looks at the results, with many high and low scoring individuals in every group (cf. also the generally large SD values), may be taken to mean that from the point of view of the kind of knowledge tested the year groups are not really groups in any but the strictly formal (administrative) sense. This is not to say that there are no differences from one student or student group to the next in overall English proficiency. For that reason it might be even more interesting to do a series of similar studies on other aspects of FL proficiency, preferably using the longitudinal design, to see what it is that our students do learn over the years. As for the present results, it is something of a surprise, if not something for the teaching staff to worry about, that so many of the 4th and 5th year people, about to start their professional careers,

did about as well, or as badly, as many of the 1st and 2nd year people on a test that drew on knowledge that may not be terribly important but is certainly not entirely useless. After all, we are basically a teacher training establishment. There is nothing to rejoice about but I was probably right when I suggested in an earlier paper [Tomaszczyk, 1987] that the language of our students' written work is sometimes not quite as satisfactory as we would like it to be not so much because they are lazy and do not bother to look things up in reference books as because most of the time they do not know they have language problems. If you do not know that you do not know something, you cannot even be expected to try to do something about it.